week_7

krishna sai surendra babu kalluri

25/07/2020

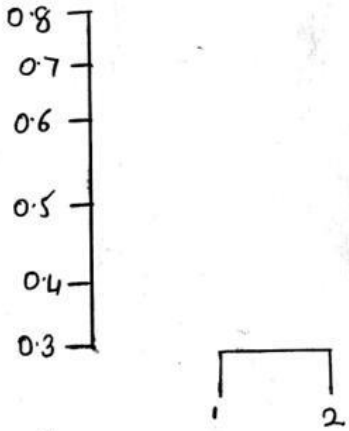
1 a)

```
knitr::include_graphics("C:\\Users\\kallu\\Documents\\images\\1a1.jpeg")
```

a) Given matrix

	P_1	P_2	P_3	P_4
P_1	0	0.3	0.4	0.7
P_2	0.3	0	0.5	0.8
P_3	0.4	0.5	0	0.45
P_4	0.7	0.8	0.45	0

Here, we can see 0.3 is minimum dissimilarity between P_1 and P_2. So, we can draw the dendogram between 1 and 2



Now, we need to recalculate the distance matrix

Distance matrix for P_3 and P_2 is $\max[dist((P_1, P_2), P_3)]$

$= \max(dist(P_2, P_3), (P_2, P_3))$

$= \max[0.4, 0.5]$

...

```
knitr::include_graphics("C:\\Users\\kallu\\Documents\\images\\1a2.jpeg")
```

$= 0.5$

Distance matrix $\max[dist(P_1, P_2), P_4]$

$= \max[dist(P_1, P_4), (P_2, P_4)]$

$= \max[0.7, 0.8]$

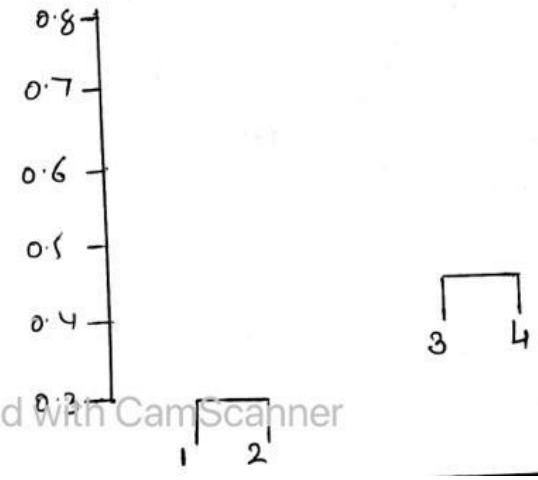
$= 0.8$

So. New matrix is,

	P_1P_2	P_3	P_4
P_1P_2	0	0.5	0.8
P_3	0.5	0	0.45
P_4	0.8	0.45	0

Now, the minimum dissimilarity is 0.45 between P_3 and P_4.

Draw the dendogram between 3 and 4.



…

```
knitr::include_graphics("C:\\Users\\kallu\\Documents\\images\\1a3.jpeg")
```

Now, we recalculate the distance matrix

Distance matrix for P_3 and P_4 is $\max[dist(P_3, P_4), P_1P_2]$

$= \max(dist(P_3, P_1P_2), (P_4, P_1P_2))$

$= \max[0.5, 0.8]$

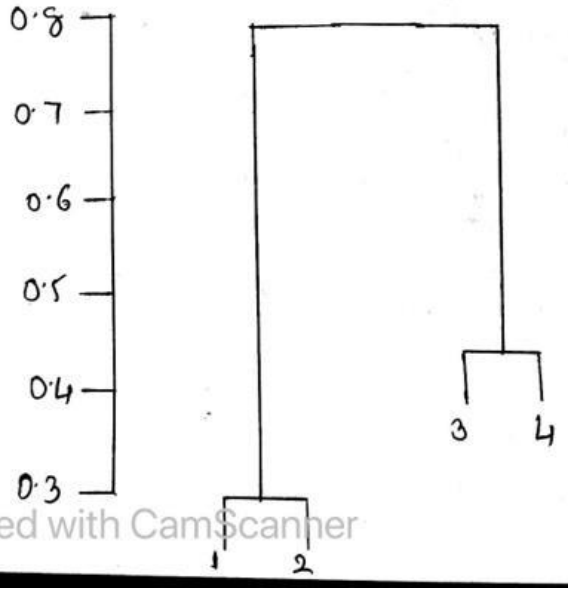
$= 0.8$

So, New matrix will be

	P_1P_2	P_3P_4
P_1P_2	0	0.8
P_3P_4	0.8	0

Here, 0.8 is the dissimilarity between P_1P_2 and P_3P_4.

Draw the dendogram



...

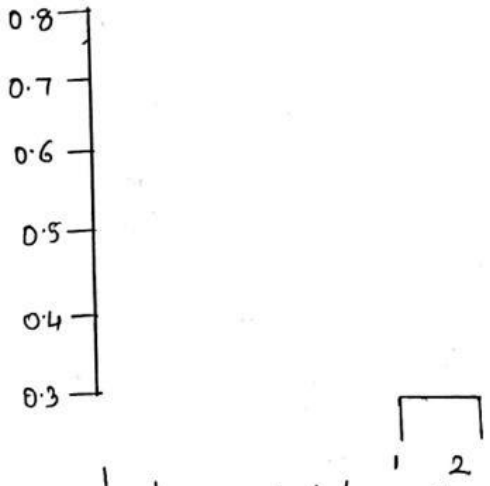
1 b)

```
knitr::include_graphics("C:\\Users\\kallu\\Documents\\images\\1b1.jpeg")
```

1) b) Given matrix Single linkage.

	P_1	P_2	P_3	P_4
P_1	0	0.3	0.4	0.7
P_2	0.3	0	0.5	0.8
P_3	0.4	0.5	0	0.45
P_4	0.7	0.8	0.45	0

Here, we can see 0.3 is minimum dissimilarity between P_1 and P_2. So, we can draw the dendogram for 1 and 2.



Now, we need to recalculate the distance matrix.

Distance matrix for P_1 and P_2 is $\min[dis(P_1, P_2), P_3]$

$= \min(dist(P_1, P_3), (P_2, P_3))$

$= \min[0.4, 0.5]$

...

```
knitr::include_graphics("C:\\Users\\kallu\\Documents\\images\\1b2.jpeg")
```

$= 0.4$

Distance matrix $\min [dist(P_1, P_2), P_4]$

$= \min [dist(P_1, P_4), (P_2, P_4)]$

$= \min [0.7, 0.8]$

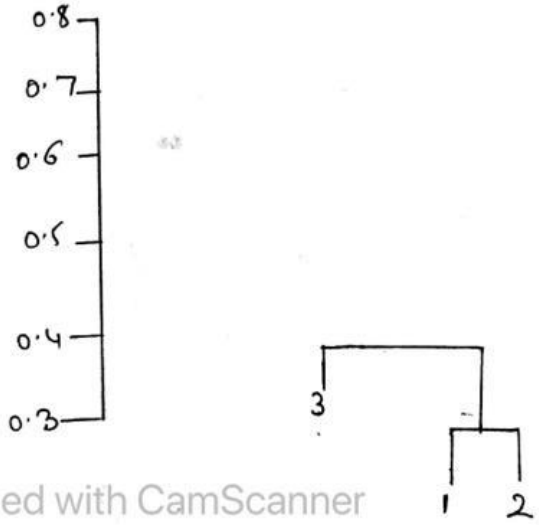
$= 0.7$

So. New matrix is

	P_1P_2	P_3	P_4
P_1P_2	0	0.4	0.7
P_3	0.4	0	0.45
P_4	0.7	0.45	0

Now, the minimum dissimilarity is 0.4 between P_1P_2 and P_3

Draw the dendogram between P_1P_2 and P_3



...

```
knitr::include_graphics("C:\\Users\\kallu\\Documents\\images\\1b3.jpeg")
```

calculate the distance matrix.

Distance matrix for P_1P_2 and P_3 is $\min[dis(P_1P_2, P_3), P_4]$

$= \min(dis(P_1P_2, P_4), (P_3, P_4))$

$= \min[0.7, 0.45]$

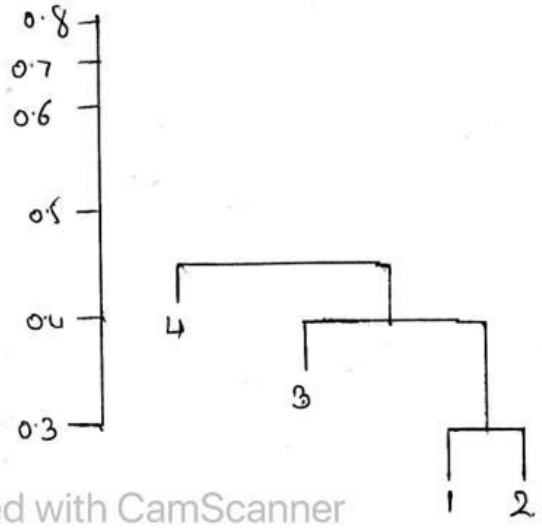
$= 0.45$

So, New matrix is

	$P_1P_2P_3$	P_4
$P_1P_2P_3$	0	0.45
P_4	0.45	0

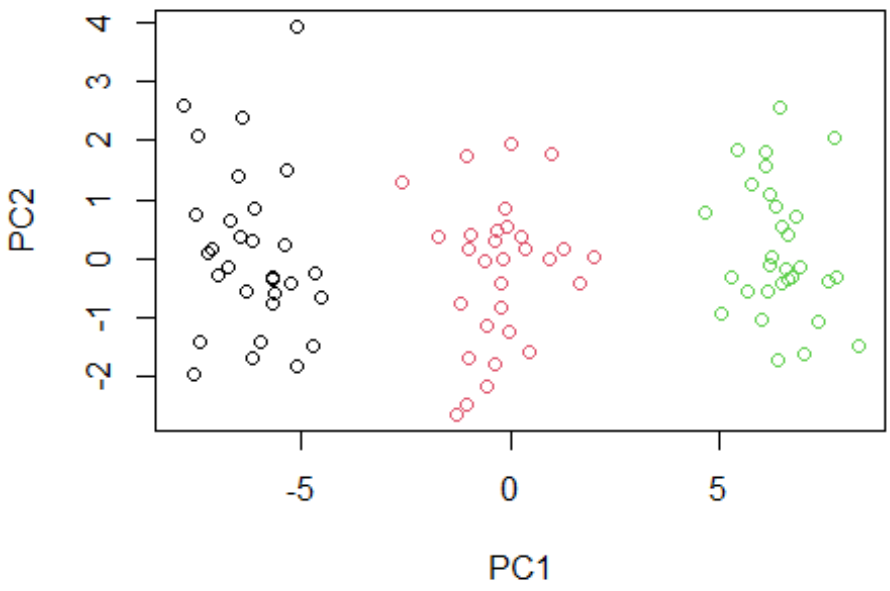
Here, 0.45 is the dissimilarity between $P_1P_2P_3$ and P_4

So, the Dendogram between $P_1P_2P_3$ and P_4 is



2 a)

```
data <- rbind(matrix(rnorm(30*10, mean = 2), nrow = 30),
matrix(rnorm(30*10, mean=4), nrow = 30),
matrix(rnorm(30*10, mean=6), nrow = 30))
data.pca = prcomp(data)$x
plot(data.pca[,1:2], col=c(rep(1,30), rep(2,30), rep(3,30)))
```




```
knitr::opts_chunk$set(echo = TRUE)
```

2 b)

```
km.out=kmeans(data,3,nstart =20)
km.out$cluster
```

```
##  [1] 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 1 1 1 1 1
1 1 1
## [39] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 2 2 2 2 2 2 2 2 2 2
2 2 2
## [77] 2 2 2 2 2 2 2 2 2 2 2 2 2
```

```
true_label = c(rep(1,30), rep(2,30), rep(3,30))
table(km.out$cluster,true_label)
```

```
##    true_label
##      1  2  3
##   1  0 30  0
##   2  0  0 30
##   3 30  0  0
```

```
knitr::opts_chunk$set(echo = TRUE)
```

The above code we performed k-means clustering with k =3 variable and we got the true and predicted labels.Then we compared both the labels and got to know that all of them are perfectly clustered .

2 c)

```
knitr::opts_chunk$set(echo = TRUE)
km.out=kmeans(scale(data),3,nstart =20)
km.out$cluster
```

```
##  [1] 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 3 3 3 3 3
3 3 3
## [39] 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 1 1 1 1 1 1 1 1 1 1 1
1 1 1
## [77] 1 1 1 1 1 1 1 1 1 1 1 1 1 1
```

```
true_label = c(rep(1,30), rep(2,30), rep(3,30))
table(km.out$cluster,true_label)
```

```
##    true_label
##      1  2  3
##   1  0  0 30
##   2 30  0  0
##   3  0 30  0
```

The above code has performed with k cluster using scale function . The results are same as the above,Here the scaling function does not have effect on results it remains to be same .

same as above results ,scaling doesnot change results

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com.

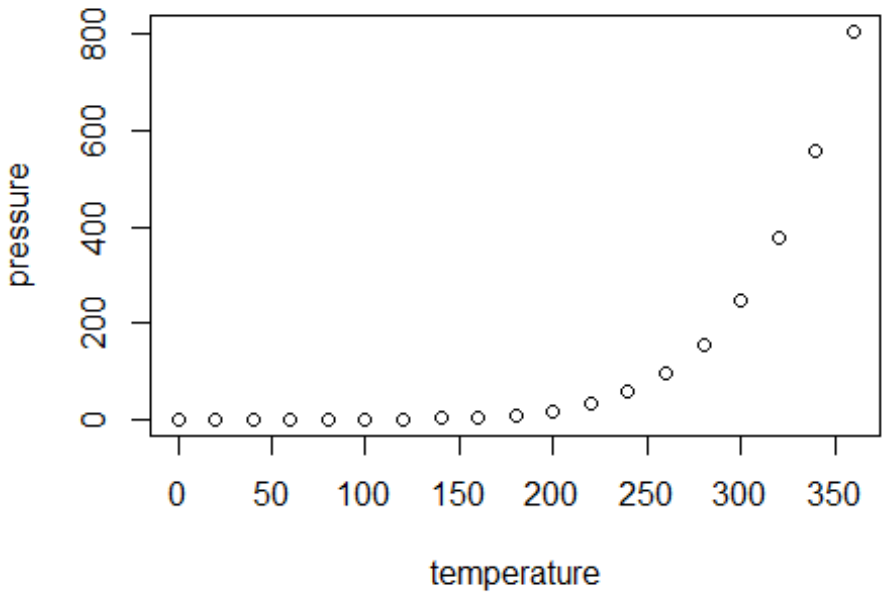
When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
summary(cars)
```

```
##      speed           dist
##  Min.   : 4.0   Min.   :  2.00
##  1st Qu.:12.0   1st Qu.: 26.00
##  Median :15.0   Median : 36.00
##  Mean   :15.4   Mean   : 42.98
##  3rd Qu.:19.0   3rd Qu.: 56.00
##  Max.   :25.0   Max.   :120.00
```

Including Plots

You can also embed plots, for example:



Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.